

# REMOVAL OF CORRELATED NOISE BY MODELING SPATIAL CORRELATIONS AND INTERSCALE DEPENDENCIES IN THE COMPLEX WAVELET DOMAIN

Bart Goossens, Aleksandra Pižurica\*, Wilfried Philips

Ghent University - TELIN - IPI - IBBT  
Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium  
bart.goossens@telin.ugent.be

## ABSTRACT

We develop a new vector-based shrinkage rule, based on the concept of "signal of interest", for the removal of correlated noise. The multivariate Bessel K Form density is used for modeling the spatial correlations between complex wavelet coefficients. The interscale dependencies between the coefficients are captured using a Hidden Markov Tree model. The combined spatial and interscale model gives improvements over recently proposed Hidden Markov Models for white noise. The results show that correlated noise is suppressed well while image details are being preserved.

**Index Terms**— Image restoration, Gaussian noise, correlated noise, Hidden Markov models

## I. INTRODUCTION

Digital cameras often produce images corrupted with noise, originating from analogue components (sensors and amplifiers) in bad lightening conditions, or at high ISO settings. By colour demosaicing and digital post-processing operations in the camera, the noise becomes correlated. During the last two decades, multiresolution concepts like wavelets have become a powerful tool for the removal of noise, and many new techniques have emerged, e.g. [1]–[6]. These methods have concentrated on white (uncorrelated) noise, resulting in simple and elegant pointwise shrinkage rules. However, when applied on images with correlated noise, these methods introduce too much smoothing, or do not remove all of the noise.

In [7], [8], a technique is proposed that deals with additive stationary correlated noise, by modeling the noise-free wavelet coefficients using a multivariate Gaussian Scale Mixture (GSM) and by treating the noise as multivariate Gaussian. A vector-based shrinkage rule, is applied on local windows of wavelet coefficients, and successfully removes the noise while maintaining image details.

Recently proposed Hidden Markov Tree (HMT) Models capture the interscale dependencies by modeling the Markovian dependencies between coefficients along the wavelet

tree structure, e.g. [2]–[4]. In [4], the spatial context of wavelet coefficients is also incorporated into the HMT model, by calculating the average energy of the neighbouring coefficients of the given coefficient.

In this paper, we develop a denoising method for the removal of correlated noise on images. We model the local spatial correlations between noise-free wavelet coefficients using the *Multivariate Bessel K Form (MBKF) prior* (Section II). We extend the method of [6] to allow multivariate priors (Section III). The notion of "signal of interest" from [6] allows us to distinguish smooth areas from edges/textures, from the observed noisy coefficients. In Section IV we build a HMT model upon this concept, resulting in a simple and efficient two state Markov model, that captures the propagation of relevant information along the wavelet tree. We discuss some implementation aspects in Section V and give visual results in Section VI.

## II. BESSEL K FORM SPATIAL PRIOR

Due to the linearity of the wavelet transform, we have the following relationship between the noise-free coefficients  $\mathbf{x}_j$ , the noise  $\mathbf{w}_j$  and the observed noisy coefficients  $\mathbf{y}_j$  on a given scale and orientation:

$$\mathbf{y}_j = \mathbf{x}_j + \mathbf{w}_j \quad (1)$$

where a one-dimensional index  $j$  denotes the spatial position (like raster scanning). The vectors  $\mathbf{x}_j$ ,  $\mathbf{w}_j$  and  $\mathbf{y}_j$  are formed by extracting wavelet coefficients in a local  $\sqrt{d} \times \sqrt{d}$  window at position  $j$ . This way, the spatial correlations of both the signal and the noise are modeled inside the local window.  $\mathbf{w}_j$  is Gaussian noise  $N(\mathbf{0}, \Sigma_w)$ .  $\mathbf{x}_j$  is modeled using the elliptically contoured *Multivariate Bessel K Form Density* (MBKF) from [9] with pdf<sup>1</sup>

$$f(\mathbf{x}) = \frac{2(2\pi)^{-d/2}}{\Gamma(\tau)|\Sigma_x|^{1/2}} \left( \frac{Q(\mathbf{x})}{\sqrt{2}} \right)^{\tau-d/2} K_{\tau-d/2}(\sqrt{2}Q(\mathbf{x})) \quad (2)$$

<sup>1</sup>In [9] this density is called "Generalized Laplace". To avoid confusion with the common definition of the Generalized Laplace from [1], [6], we use the name *Bessel K Form*, as in [10], [11] for the univariate counterpart.

\*A. Pižurica is a postdoctoral researcher of the Fund for the Scientific Research in Flanders (FWO) Belgium

where  $Q(\mathbf{x}) = \sqrt{\mathbf{x}^T \boldsymbol{\Sigma}_x^{-1} \mathbf{x}}$ ,  $K_m(u)$  is the modified Bessel function of the second kind and order  $m$  (see [9], [12]), and  $\Gamma(\tau) = \int_0^\infty z^{\tau-1} e^{-z} dz$  is the Gamma function. For  $\tau = 1$ , (2) simplifies to the Multivariate Laplace distribution, used in [12]. For the limit  $\tau \rightarrow \infty$ , (2) approaches the Gaussian distribution  $N(\mathbf{0}, \boldsymbol{\Sigma}_x)$ . The parameter range  $\tau = ]0, 1]$  is of special interest, because it allows us to model the highly kurtotic non-Gaussian behaviour of wavelet coefficients. The kurtosis is  $\gamma_2 = 3 + 3/\tau$ , which approaches  $+\infty$  for small  $\tau > 0$ . Furthermore, the MBKF admits the following *Gaussian Scale Mixture* (GSM) representation [9]:

$$\mathbf{x} \stackrel{d}{=} z^{1/2} \mathbf{u} \quad (3)$$

where " $\stackrel{d}{=}$ " means *equality in distribution*,  $\mathbf{u}$  is *Gaussian*  $N(\mathbf{0}, \boldsymbol{\Sigma}_x)$  and  $z$  is *Gamma* distributed  $\Gamma(\alpha = \tau, \beta = 1)$ :

$$f(z) = \frac{1}{\Gamma(\tau)} z^{\tau-1} e^{-z} \quad (4)$$

Finally, we estimate parameters  $\boldsymbol{\Sigma}_y$  using maximum likelihood estimation (MLE),  $\tau$  by matching the second and fourth order cumulants of marginal distribution with the observation, as explained in [11] and  $\boldsymbol{\Sigma}_x$  using  $\hat{\boldsymbol{\Sigma}}_x = \hat{\tau}^{-1}(\hat{\boldsymbol{\Sigma}}_y - \boldsymbol{\Sigma}_w)$ . We further assume that  $\boldsymbol{\Sigma}_w$  is known.

### III. "SIGNAL OF INTEREST" BASED SHRINKAGE RULE

#### III-A. Definition of "signal of interest"

Noise-free wavelet coefficients can be categorized into:

- Smooth areas, containing *small* coefficient magnitudes
- Edges and textures, exhibiting *large* coefficient magnitudes (we will call this *signal of interest*)

One way to distinguish both groups, is using a significance map, e.g. [5], [6] for  $d = 1$ :

$$S(x) = \mathbb{I}(|x|/\sigma \geq T) \quad (5)$$

where  $T$  is a threshold,  $\sigma$  is the noise standard deviation and  $\mathbb{I}(x)$  is the indicator function. We propose the following generalization for  $d > 1$ :

$$S(\mathbf{x}) = \mathbb{I}(\|\mathbf{C}_w \mathbf{x}\| \geq T) \quad (6)$$

where  $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}$  is the *Frobenius*-norm of  $\mathbf{x}$  and  $\mathbf{C}_w = \boldsymbol{\Sigma}_w^{-1/2}$ . A geometrical explanation is that  $\|\mathbf{C}_w \mathbf{x}\| = T$  represents the equation of an *ellipsoid* in the  $d$ -dimensional Euclidean space ( $\boldsymbol{\Sigma}_w$  is positive definite), and  $S(\mathbf{x})$  tests whether  $\mathbf{x}$  is inside or outside the ellipsoid. In the following, we will denote  $S(\mathbf{x}) = i$  as hypothesis  $H_i$ ,  $i = 0, 1$ .

#### III-B. Bayesian shrinkage rule

We now consider the following simple shrinkage rule:

$$\hat{\mathbf{x}}_j = P(H_1 | \mathbf{y}_j) \mathbf{y}_j \quad (7)$$

The heuristic motivation is as follows: each coefficient vector is scaled with the probability that it represents signal of

interest. We note that for the assumed model,  $\mathbf{y}$  follows an infinite *Gaussian Mixture* (GM), since  $\mathbf{x}$  follows a MBKF distribution (see Section II) and  $\mathbf{w}$  is Gaussian distributed. By imposing that  $f(\mathbf{x}|H_0)$  is an infinite GM (and as a consequence also  $f(\mathbf{y}|H_0)$ ), we can avoid the (excessive) numerical calculation of the convolution  $f(\mathbf{y}|H_0) = f(\mathbf{x}|H_0) \star f(\mathbf{w})$ . This condition is satisfied by:

$$f(H_0 | \mathbf{x}) = \exp\left(-\frac{1}{2T^2} \|\mathbf{C}_w \mathbf{x}\|^2\right) \quad (8)$$

We can expand (7) to:

$$\hat{\mathbf{x}}_j = \mathbf{y}_j - \left(\int_0^{+\infty} f(z) P(H_0 | z) \frac{f(\mathbf{y}_j | H_0, z)}{f(\mathbf{y}_j)} dz\right) \mathbf{y}_j \quad (9)$$

If we denote the normal density evaluated in  $\mathbf{y}$  as  $N(\mathbf{y}; \mathbf{0}, \boldsymbol{\Sigma})$ , and using (8) we can write:

$$f(\mathbf{y} | z) = N(\mathbf{y}; \mathbf{0}, z\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_w)$$

$$f(\mathbf{y} | H_0, z) = N\left(\mathbf{y}; \mathbf{0}, ((z\boldsymbol{\Sigma}_x)^{-1} + (T^2\boldsymbol{\Sigma}_w)^{-1})^{-1} + \boldsymbol{\Sigma}_w\right)$$

Using the above equations, it is straightforward to show that:

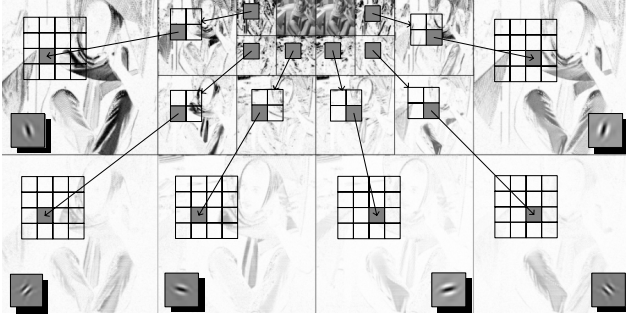
$$P(H_0 | z) = \left(\frac{|(z\boldsymbol{\Sigma}_x)^{-1}|}{|(z\boldsymbol{\Sigma}_x)^{-1} + (T^2\boldsymbol{\Sigma}_w)^{-1}|}\right)^{1/2} \quad (10)$$

which allows us to calculate the probability that the signal of interest is present/absent on the considered band:

$$P(H_0) = 1 - P(H_1) = \int_0^\infty P(H_0 | z) f(z) dz \quad (11)$$

### IV. HIDDEN MARKOV TREE MODEL FOR INTERSCALE DEPENDENCIES

A detailed explanation of Hidden Markov Tree models can be found elsewhere (e.g. [2], [13]). Here we will only discuss the application to our problem. First, we want to characterize the interscale dependencies of wavelet coefficients, along a given wavelet tree (see Fig. 1). We could accomplish this by simply modeling the joint probability density of the whole wavelet tree, and by applying a vector-based shrinkage rule like (7). However, in many cases, the multivariate BKF density (or any other GSM model) does not provide a good fitting to the multiscale data, because the joint probability histograms of coefficients are generally not elliptically contoured anymore. Working with high-dimensional probability densities also requires larger covariance matrices (with size  $d \times d$ ), and more parameters to estimate. At a given point, it becomes impossible to find good estimates because 1) estimation becomes less reliable because there is not enough data and 2) overfitting of the data becomes a real problem. Therefore, we want to track only the *key*-dependencies between the wavelet coefficients, i.e. intrascale (spatially) and interscale. The HMT models Markovian dependencies between wavelet coefficients in the wavelet tree, thereby reducing the number of model parameters. With each wavelet coefficient, a hidden node



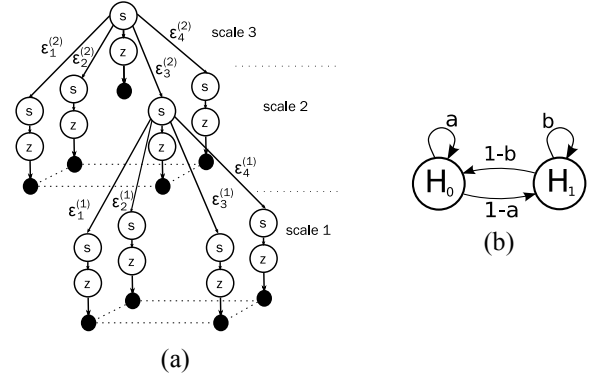
**Fig. 1.** Independent HMT models associated with the 6 orientations of the Dual Tree Complex Wavelet Transform (DTCWT) of [15] (background: complex coefficient magnitudes, gray boxes: real parts of the corresponding complex wavelets)

is associated. Every hidden node is considered to be in a set of  $M$  distinct states. The hidden nodes are connected in a directed tree starting from the root node, and it is assumed that the state of each node only depends on the state of the parent node (Markov property, see Fig. 2a). Due to computational constraints, it is preferable to limit the number of states. We use the significance measures  $S(\mathbf{x})$  as hidden nodes for the Markov model, following Malfait's idea from [14] for spatial Markov Random Fields, also used in [5]. This results in a two state HMT model with two hidden layers (Fig. 2), parametrized by  $\Theta = \{\Sigma_{x,n}, \tau_n, \alpha_{n,i}, \epsilon_n^{(k)}\}$  where  $n = 1, \dots, N$  denotes the scale,  $\alpha_{n,i} = P(H_i^{(n)})$ , the probability that the signal of interest is absent/present on scale  $n$  and  $\epsilon_n^{(k)}$  is the state transition probability matrix for scale  $n$ .  $k = 1, \dots, 4$  represents the child index of the hidden node on scale  $n$  (see Fig. 2a). By assigning different transition probability matrices to each child node of a given node, we can better adapt the HMT model to the spatial context of the finer scale. The parameters  $\Sigma_{x,n}$  and  $\tau_n$  are estimated once for every scale, while  $\alpha_{n,i}, \epsilon_n^{(k)}$  are estimated recursively using the *Baum-Welch* algorithm [2], [13]. We found experimentally that for many images,  $(\epsilon_n^{(k)})_{1,1} \approx 1$  and  $(\epsilon_n^{(k)})_{1,2} \approx 0$ . This means that, once a coefficient is *insignificant*, it remains *insignificant* when we go to a finer scale.

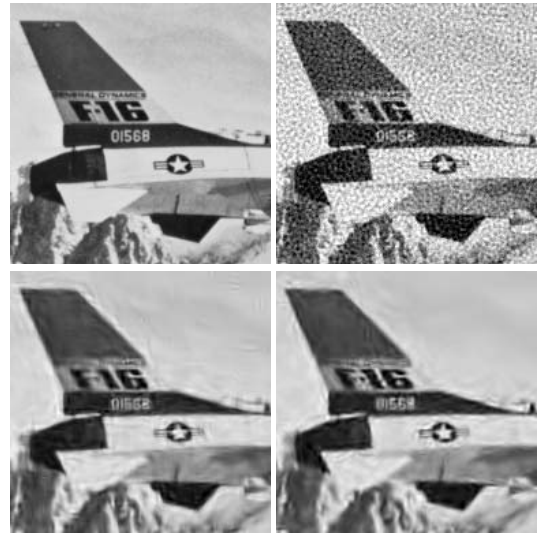
## V. IMPLEMENTATION ASPECTS

Eq. (9) has no closed-form analytical expression and requires integration over an infinite interval. However, by noting that  $f(z)$  decays fast, we can choose an upper bound for the integration and calculate the integral numerically. We found experimentally that the points  $z_l = \exp(-3 + 6l/L)$ ,  $l = 1, \dots, L$  provide a good approximation, even with  $L$  small (e.g.  $L = 4$ ).

To avoid numerical underflow in the *upward/downward* algorithm, used for likelihood computation in the HMT model, a *scaling procedure* is required [13]. For our spatial prior, we found that for large  $d$ , underflow already arises when evaluating normal densities  $f(\mathbf{y}|z)$  and  $f(\mathbf{y}|z, H_0)$ ,



**Fig. 2.** (a) proposed HMT structure (s-nodes represent the significance map, z-nodes represent the local variance, black nodes are wavelet coefficients) (b) Two-state model

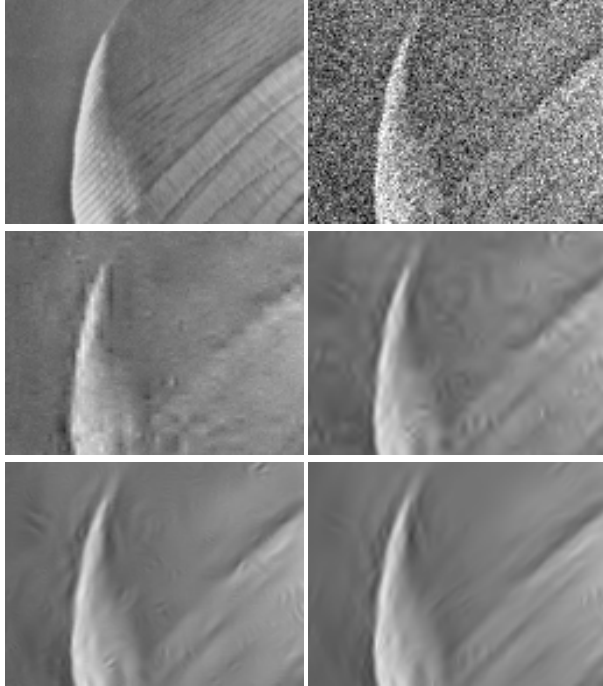


**Fig. 3.** Denoising results for correlated noise, from left to right and from top to bottom: the original image (cropped out), image with bandpass filtered noise (PSNR = 20.17dB), GSM-BLS using Full Steerable Pyramids [8] (PSNR = 29.96dB), the proposed method using DTCWT (PSNR = 29.88dB)

and the scaling procedure subsequently fails, because the input probabilities are 0. We solve this problem by adding an extra scaling factor  $e^{Ad}$  to the normal density, which will be cancelled automatically later in the *upward/downward* steps.

## VI. RESULTS

The results for this paper are produced using the Dual Tree Complex Wavelet transform of [15] using 6-tap Q-shift filters. For simplicity, we assumed that the real and imaginary parts of the complex wavelet coefficients are statistically independent, resulting in 12 independent HMT models for the 6 orientations. The HMT training step is performed on the noisy image. We use overlapping  $3 \times 3$  spatial windows ( $d = 9$ ), and only estimate the central window coefficient using (7). For this window size, we optimized the threshold



**Fig. 4.** Denoising results for white noise, from left to right and from top to bottom: the original image (cropped out), the noisy image  $\sigma = 35$  (PSNR = 17.48dB), HMT method of [2] using decimated DWT (PSNR = 27.75dB), method of [6] using undecimated DWT (PSNR = 29.31dB), GSM-BLS [7] using Full Steerable Pyramids (PSNR = 30.15dB), the proposed method using DTCWT (PSNR = 30.20dB)

$T$  experimentally, and obtained  $T = 2.66$ . In Fig. 3, a result for bandpass filtered noise is given, with Power Spectral Density  $PSD(f_x, f_y) \sim \exp(-60|f_x^2 + f_y^2 - 0.05|)$ . Here  $f_x$  and  $f_y$  are respectively the horizontal and vertical frequencies. The noise covariance matrix is available to the algorithms. By exploiting the interscale dependencies, the smooth areas remain smooth after denoising and contain less wavelet artifacts. In Fig. 4, we compare the visual denoising performance for images with *white* noise (i.e.  $\Sigma_w$  is diagonal). The HMT method of [2] still leaves some noise on the image. This is mainly caused by the lack of spatial information in the estimation of the posterior state probabilities of the hidden nodes. In our approach, by incorporating spatial correlations, the denoising performance using Complex Wavelets (with redundancy 4) is raised to the same level as the Full Steerable Pyramid (with redundancy 18.67 for 8 orientations) from [7]. Finally, on a Pentium IV 2.4GHz, our C++ implementation takes 3.5 s for denoising a  $512 \times 512$  grayscale image.

## VII. CONCLUSION

A new image denoising method for the removal of correlated noise has been presented. The multivariate Bessel K Form density is capable of modeling the local spatial correlations between wavelet coefficients, and its use is practical since it has a Gaussian Scale Mixture representation where

the hidden multiplier is Gamma distributed. A significance measure allows us to quantify the relevant information in a noisy image, and a Hidden Markov Tree model built on this measure, captures the propagation of this information along the wavelet coefficient tree. The results show that, by using a combined spatial and interscale prior, image details are better reconstructed while smooth areas in the original image remain smooth after denoising.

## VIII. REFERENCES

- [1] S. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, pp. 1522–1531, 2000.
- [2] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, pp. 886–902, 1998.
- [3] J. Romberg, H. Choi, R. Baraniuk, and N. G. Kingsbury, "Hidden Markov tree modeling of complex wavelet transforms," in *Proc. IEEE Conf. on Image Process.*, 2001.
- [4] G. Fan and X. Xia, "Image denoising using local contextual hidden Markov model in the wavelet domain," *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 125–128, May 2001.
- [5] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 5, pp. 545–557, 2002.
- [6] A. Pižurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 654–665, Mar 2006.
- [7] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using Gaussian Scale Mixtures in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, pp. 1338–1351, 2003.
- [8] J. Portilla, "Full blind denoising through noise covariance estimation using Gaussian Scale Mixtures in the wavelet domain," *IEEE International Conference on Image Processing (ICIP)*, vol. 2, pp. 1217–1220, 2004.
- [9] S. Kotz, T. Kozubowski, and K. Podgorski, *The Laplace Distributions And Generalizations: A Revisit with Applications to Communications, Economics, Engineering, Finance*. Birkhäuser, 2001.
- [10] A. Srivastava, X. Liu, and U. Grenander, "Universal Analytical Forms for Modeling Image Probabilities," *IEEE Trans. Pat. Analysis and Machine Intell.*, vol. 24, no. 9, pp. 1200–1214, Sept. 2002.
- [11] J. M. Fadili and L. Boubchir, "Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities," *IEEE Trans. on Image Process.*, vol. 14, no. 2, pp. 231–240, Feb. 2005.
- [12] I. W. Selesnick, "Laplace random vectors, Gaussian noise, and the generalized incomplete Gamma function," in *Proc. IEEE Conf. on Image Process.*, 2006, pp. 2097–2100.
- [13] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [14] M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov Random Field a priori model," *IEEE Trans. Image Proc.*, vol. 6, no. 4, pp. 549–565, April 1997.
- [15] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.